AD-A063 984 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
ON THE VALIDITY OF A NONLINEAR PROGRAMMING METHOD FOR SOLVING M--ETC(U)
OCT 78 S HAN DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1891 NL

1 OF 1
ADA
063984

END
DATE
FILMED
4-79
DDC

AD A063984

(10) NA

(9) MRC Technical Summary Report, #1891

(6) ON THE VALIDITY OF A NONLINEAR PROGRAMMING METHOD FOR SOLVING MINIMAX PROBLEMS.

(10) Shih-Ping/Han

(11) Oct 78

(12) 14p.

LEVEL II

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

October 1978

(15) DAAG29-75-C-ΦΦ24

(Received August 16, 1978)

(14) MRC-TSR-1891

Approved for public release
Distribution unlimited

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

221 200

79 01 30 037

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ON THE VALIDITY OF A NONLINEAR PROGRAMMING METHOD FOR SOLVING MINIMAX PROBLEMS

Shih-Ping Han

Technical Summary Report #1891
October 1978

ABSTRACT

We consider the minimization of a function which is the maximum of a finite number of smooth but nonlinear functions. It is well-known that the minimax problem of this type connects naturally to a nonlinear program. Through this connection the effective quasi-Newton method becomes applicable. We show that this approach is valid and the resulting method has global convergence properties.

AMS(MOS) Subject Classification - 65K05, 90C30

Key Words - Minimax, Nonlinear programming, Quasi-Newton method

Work Unit Number 5 - Mathematical Programming and Operations Research

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

SIGNIFICANCE AND EXPLANATION

Minimax is an important principle in optimal selection of parameters in the processing of empirical data and abounds with applications in Economics, Statistics, Engineering and many other areas. The type of minimax problem considered here is to minimize a function which is the maximum of a finite number of smooth but nonlinear functions. Because the maximum function is usually not smooth, most optimization techniques are not suitable for handling it. However, by transforming the problem into a equivalent nonlinear program the efficient quasi-Newton method becomes applicable. This work shows that this approach is valid and effective.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ON THE VALIDITY OF A NONLINEAR PROGRAMMING
METHOD FOR SOLVING MINIMAX PROBLEMS

Shih-Ping Han

1. Introduction

We are dealing the following minimax problem

$$\text{(1.1)} \qquad \underset{x \in R^n}{\text{minimize}} \ \varphi(x)$$

where $\varphi(x) = \max_{i=1,\ldots,m} \{f_i(x)\}$ and f_i's are continuously differentiable real-valued functions defined on R^n. The function φ is usually not differentiable at a solution point; therefore, most unconstrained optimization methods are no longer appropriate for handling it. But, Problem (1.1) can be put into the following equivalent nonlinear programming form

$$\text{(1.2)} \qquad \begin{array}{ll} \underset{(x,\eta) \in R^{n+1}}{\text{minimize}} & \eta \\ \text{s.t.} & f_i(x) \leqq \eta \quad i = 1,\ldots,m . \end{array}$$

Hence, for solving the minimax problem (1.1) it seems feasible to use an effective nonlinear programming method to solve (1.2). A purpose of this paper is to demonstrate how the successful quasi-Newton method described in [3, 4, 8] can be so used. The special structure that Problem (1.2) is linear in the variable η should be exploited. But, with this being done, the global convergence theorems in [3] can no longer apply here. We show in this paper that the resulting method is still a valid one and global convergence is still achievable.

We describe the method in Section 2. Section 3 is devoted to the justification of our approach. The global convergence theorem for the method is given in Section 4.

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

2. The Method

We are content with finding a stationary point of Problem (1.1), by which it is meant a point, $\bar{x}$ say, in R^n that satisfies the condition

$$\min \{\varphi'(\bar{x};d) : \|d\|_2 = 1\} \leqq 0 , \tag{2.1}$$

where $\varphi'(\bar{x};d)$ is the directional derivative of φ at $\bar{x}$ in the direction d. Clearly, Condition (2.1) is a necessary condition for the point $\bar{x}$ to be a solution of (1.1) and it will be reduced to the condition $\nabla\varphi(\bar{x}) = 0$ when $m = 1$. Let $I(x) = \{i : f_i(x) = \varphi(x)\}$ and let f_i be called active at x if $i \in I(x)$. It is known [2, for instance] that $\bar{x}$ is stationary if and only if there exists an m-vector $\bar{v}$ such that

(2.2)

$$\text{(a)} \quad \sum_{i=1}^{m} \bar{v}_i \nabla f_i(x) = 0 ;$$

$$\text{(b)} \quad \sum_{i=1}^{m} \bar{v}_i = 1 ;$$

$$\text{(c)} \quad \bar{v} \geqq 0 ;$$

$$\text{(d)} \quad \bar{v}_i = 0 \quad \text{if} \quad i \notin I(\bar{x})$$

Notice that Condition (2.2) is just the Karush-Kuhn-Tucker condition of the nonlinear programming problem (1.2).

The proposed method is an iterative process. At the k-th iteration we have an estimate x_k of a solution and have also a scaler η_k which is a predicted optimal value of the objective function φ. To construct new estimates x_{k+1} and η_{k+1} we solve the following quadratic program

$$\underset{(d,\delta)\in R^{n+1}}{\text{minimize}} \quad \frac{1}{2} d^T B_k d + \delta \tag{2.3}$$

$$\text{s.t.} \quad f_i(x_k) + \nabla f_i(x_k)^T d \leqq \eta_k + \delta \qquad i = 1,\ldots,m .$$

Here, B_k is a positive definite $n\times n$ matrix, preferably a good approximation to the Hessian $\sum_{i=1}^{m} \bar{v}_i \nabla^2 f_i(\bar{x})$ of the Lagrangian of (1.2) and updated by Powell's scheme [9]

Let (d_k, δ_k) be a solution of (2.3). Then we set $(x_{k+1}, \eta_{k+1}) = (x_k, \eta_k) + \lambda_k(d_k, \delta_k)$, where λ_k is the stepsize determined by doing an exact line-search in the direction (d_k, δ_k) on the function θ defined by

$$\theta(x,\eta) = \eta + \sum_{i=1}^{m} \max\{f_i(x) - \eta, 0\} ; \tag{2.4}$$

that is,

$$\theta(x_{k+1}, \eta_{k+1}) = \min_{0 \leqq \lambda \leqq 1} \theta(x_k + \lambda d_k, \eta_k + \lambda \delta_k) .$$

It is merely for simplicity that we consider the exact line-search here. An analysis of some inexact line-search is possible and will be very similar to the one given in [5], where the determination of the stepsite λ_k is based on the objective function φ instead of θ. There are some advantages to use the function θ because it takes into consideration some inactive functions f_i, while the function φ gives bias completely to the active ones.

The method described above is essentially an application of the method in [3] to Problem (1.2) with its special structure being taken into account. The problem considered there is the general nonlinear programming problem

$$\begin{aligned} &\min \quad g(x) \\ &\text{s.t.} \quad f_i(x) \leqq 0 \quad i = 1,\ldots,m \end{aligned} \tag{2.4}$$

and the subproblem to be solved in each iteration is the quadratic program

$$\begin{aligned} &\min_{d \in R^n} \quad \nabla g(x_k)^T d + \frac{1}{2} d^T A_k d \\ &\text{s.t.} \quad f_i(x_k) + \nabla f_i(x_k)^T d \leqq 0 \quad i = 1,\ldots,m . \end{aligned} \tag{2.5}$$

The stepsizes are determined by the exact penalty function

$$p(x,\alpha) = g(x) + \alpha \sum_{i=1}^{m} \max\{f_i(x), 0\} .$$

When $\beta x^T x \geqq x^T A_k x \geqq \gamma x^T x$ for some positive numbers β and γ and for each k and x, and when the Lagrange multipliers of (2.5) are uniformly bounded by α in the ∞-norm then it is shown in [3] that any accumulation point of the generated points

$\{x_k\}$ is a Kuhn-Tucker point of (2.4). Because Problem (1.2) is linear in η, hence the quadratic program (2.3) should preserve this property. Though we usually require the matrix B_k to be positive definite, Problem (2.3) is no longer a positive definite quadratic program in the space R^{n+1}. This makes the global convergence theorems in [3] not applicable here. Some justification of our approach becomes necessary and will be given in the following sections.

3. Validity of the method

In this section we justify our approach. For simplicity we drop the index k from (2.3) and for any point (x,η) in R^{n+1} and any $n\times n$ matrix B we consider the following quadratic program

$$
\text{(3.1)} \qquad \begin{aligned} &\underset{(d,\delta)\in R^{n+1}}{\text{minimize}} \ \frac{1}{2} d^T B d + \delta \\ &\text{s.t.} \quad f_i(x) + \nabla f_i(x)^T d \leqq \eta + \delta \qquad i = 1,\ldots,m \,. \end{aligned}
$$

Because the constraints are linear any solution of (3.1), $(\bar{d},\bar{\delta})$ say, is a Karush-Kuhn-Tucker point of (3.1) and, hence, there exists an m-vector $\bar{v}$ such that

$$
\text{(3.2)} \qquad \begin{aligned} &\text{(a)} \quad B\bar{d} + \sum_{i=1}^{m} \bar{v}_i \nabla f_i(x) = 0 \,; \\ &\text{(b)} \quad \sum_{i=1}^{m} \bar{v}_i = 1 \,; \\ &\text{(c)} \quad f_i(x) + \nabla f_i(x)^T \bar{d} \leqq \eta + \bar{\delta}, \quad i = 1,\ldots,m \,; \\ &\text{(d)} \quad \bar{v} \geqq 0 \,; \\ &\text{(e)} \quad \bar{v}_i = 0 \ \text{ if } \ f_i(x) + \nabla f_i(x)^T \bar{d} < \eta + \bar{\delta} \,. \end{aligned}
$$

We first show that the method is well-defined in the sense that a search direction can always be uniquely determined in each iteration.

<u>Theorem 3.1</u>. Let (x,η) be any point in R^{n+1} and let B be a positive definite $n\times n$ matrix. Then there exists a unique solution $(\bar{d},\bar{\delta})$ of the quadratic program (3.1). Furthermore, x is a stationary point of (1.1) if and only if $\bar{d} = 0$ and $\bar{\delta} = \varphi(x) - \eta$.

<u>Proof</u>. To show the existence of a solution for (3.1) we first note that its feasible region is obviously non-empty. Because of the convexity we only need to show that the objective function is bounded below in the feasible region. This can be done by considering the dual problem

$$\max_{v \in R^m} \; f(x)^T v - \frac{1}{2} v^T \nabla f(x) B^{-1} \nabla f(x)^T v$$

$$\text{s.t.} \quad \sum_{i=1}^{m} v_i = 1 \,,$$

$$v \geqq 0 \,.$$

Clearly, the dual problem is also feasible. Therefore, it follows from the duality theorem [7] that the optimal value of (3.1) is finite and, hence, a solution exists.

Let $(\bar{d},\bar{\delta})$ be a solution of (3.1). We want to show that $(\bar{d},\bar{\delta})$ is the only one. From a result in [6, Corollary 3.6] the solution $(\bar{d},\bar{\delta})$ is unique if $d^T B d > 0$ for any non-zero vector (d,δ) in R^{n+1} satisfying

$$\text{(3.3)} \qquad \text{(a)} \quad \bar{d}^T d + \delta \leqq 0 \,,$$

$$\text{(b)} \quad \nabla f_i(x)^T d \leqq \delta \quad \text{for each } i \in J \,,$$

where $J = \{j : f_i(x) + \nabla f_i(x)^T \bar{d} = \eta + \bar{\delta}\}$. Because B is positive definite we only need to show that if (d,δ) satisfies (3.3) and $d = 0$ then δ must also be zero. Note that it follows from (3.2.b) and (3.2.e) that the index set J can not be empty. Therefore, if $d = 0$ and if (d,δ) satisfies (3.3.b) for some i in J then $\delta \geqq 0$. But, from (3.3.a) and $d = 0$ we also have $\delta \leqq 0$. Hence, $\delta = 0$ and the uniqueness of the solution is proven.

The second part of the theorem follows straightforwardly from (3.2) and the uniqueness of the solution. Q.E.D.

It may be worthwhile noticing that the solution vector $\bar{d}$ is independent of the given value η. Therefore, a bad estimate in η should usually not spoil a good estimate in x.

The search direction (d,δ) is not only well-defined but also useful because it is descent for our optimality function θ. Before giving this result we note here that the symbol $\theta'(x,\eta;d,\delta)$ denotes the directional derivative of θ at (x,η) in the direction (d,δ).

Theorem 3.2. Let (x,η) be a given point in R^{n+1} and B be a positive definite and symmetric $n\times n$ matrix. Let $(\bar{d},\bar{\delta})$ be the unique solution of (3.1). Then $\theta'(x,\eta;\bar{d},\bar{\delta}) \leq -\bar{d}^T B\bar{d}$.

Proof. Let $I_+ = \{i : f_i(x) > \eta\}$, $I_0 = \{i : f_i(x) = \eta\}$, and $I_- = \{i : f_i(x) < \eta\}$. Then from a result of Danskin [1] we have

$$\theta'(x,\eta;\bar{d},\bar{\delta}) = \bar{\delta} + \sum_{I_+} (\nabla f_i(x)^T\bar{d} - \bar{\delta}) + \sum_{I_0} \max\{\nabla f_i(x)^T\bar{d} - \bar{\delta},0\} .$$

Because $(\bar{d},\bar{\delta})$ is also a Karush-Kuhn-Tucker point of (3.1) there exists an m-vector $\bar{v}$ such that (3.2) holds. It follows from (3.2.c) that if $i \in I_0$ then $\max\{\nabla f_i(x)^T\bar{d} - \bar{\delta},0\} = 0$. It also follows from (3.2) that

$$\begin{aligned}\theta'(x,\eta;\bar{d},\bar{\delta}) &= -\bar{d}^T B\bar{d} - \sum_{i=1}^{m} \bar{v}_i\nabla f_i(x)^T\bar{d} + \bar{\delta} + \sum_{I_+} (\nabla f_i(x)^T\bar{d} - \bar{\delta}) \\ &\leq -\bar{d}^T B\bar{d} - \sum_{i=1}^{m} \bar{v}_i(\bar{\delta} + \eta - f_i(x)) + \bar{\delta} + \sum_{I_+} (\eta - f_i(x)) \\ &\leq -\bar{d}^T B\bar{d} + \sum_{I_+} (1 - \bar{v}_i)(\eta - f_i(x)) \leq -\bar{d}^T B\bar{d} .\end{aligned}$$

Q.E.D.

The following corollary justifies the use of function θ as an optimality function for solving the minimax problem (1.1).

Corollary 3.3. If (x,η) is a local minimum point of function θ then x is a stationary point of Problem (1.1).

Proof. Consider the quadratic program (3.1) with B being any positive definite and symmetric matrix. Let $(\bar{d},\bar{\delta})$ be its solution. The vector $\bar{d}$ must be zero; otherwise, $\theta'(x,\eta;\bar{d},\bar{\delta}) < 0$ which contradicts that (x,η) is a local minimum. Hence, the result follows immediately from Theorem 3.1. Q.E.D.

We also observe that both Problem (1.2) and (3.1) satisfy the Arrow-Hurwicz-Uzawa constraint qualification [see 7, for instance]. From a result in [10; Theorems 1 and 3] the feasible regions of (1.2) and (3.1) are stable when they are subjected to small perturbations. This property is very desirable and makes our approach very useful pratically.

4. Global convergence

We study in this section the global convergence properties of the method. Here we need a very useful result of Robinson [11] on the stability of quadratic programs. Actually, Robinson considers the stability of a very general class of problems, called generalized equations by him. The following lemma is a straightforward consequence of his Theorem 2.

Lemma 4.1. Let (x,η) be a point in R^{n+1} and let B be a positive definite $n\times n$ matrix, and let $(\bar{d},\bar{\delta})$ be the unique solution of (3.1). Then there exist constants λ and ε such that for each $n\times n$ matrix B' and each (x',η') in R^{n+1} with

$$\varepsilon' := \max\{\|B' - B\|_2, \|(x',\eta') - (x,\eta)\|_2\} < \varepsilon$$

the quadratic program

$$\min_{(d,\delta)\in R^{n=1}} \frac{1}{2} d^T B' d + \delta$$

$$\text{s.t.} \quad \nabla f_i(x') + f_i(x')^T d \leqq \eta' + \delta \qquad i = 1,\dots,m$$

has a unique solution (d',δ') and

$$\|(\bar{d}',\bar{\delta}') - (d',\delta')\|_2 \leqq \lambda\varepsilon'(1 - \lambda\varepsilon')^{-1}(1 + \|(\bar{d},\bar{\delta})\|_2).$$

We now give the global convergence theorem below.

Theorem 4.2. Let $\{B_k\}$ be a sequence of $n\times n$ symmetric matrices satisfying that for some positive numbers α and β and for each k

$$\beta x^T x \leqq x^T B_k x \leqq \alpha x^T x \qquad \text{for any } x \text{ in } R^n.$$

Let $\{(x_k,\eta_k)\}$ be a sequence of points in R^{n+1} generated by the method from any given starting point (x_0,η_0) and let $(\bar{x},\bar{\eta})$ be any accumulation point of this sequence. Then $\bar{x}$ is a stationary point of the minimax problem (1.1).

Proof: Without loss of generality we may assume that $(x_k,\eta_k) \to (\bar{x},\bar{\eta})$. By passing to a subsequence, if necessary, we have a positive definite and symmetric matrix $\bar{B}$ such that $B_k \to \bar{B}$. Consider the quadratic program

$$\min_{(d,\delta)\in R^{n+1}} \quad \frac{1}{2} d^T Bd + \delta$$

$$\text{s.t.} \quad f_i(\bar{x}) + \nabla f_i(\bar{x})^T d \leqq \bar{\eta} + \delta \qquad i = 1,\ldots,m\ .$$

By Theorem 3.1 the quadratic program has a unique solution, $(\bar{d},\bar{\delta})$ say. If $\bar{d} = 0$ then by Theorem 3.1 again $\bar{x}$ is a stationary point of (1.1) and the proof is done. If $\bar{d} \neq 0$ we will deduce a contradiction.

Define a point $(\hat{x},\hat{\eta})$ by $(\hat{x},\hat{\eta}) = (\bar{x},\bar{\eta}) + \bar{\lambda}(\bar{d},\bar{\delta})$ where $0 \leqq \bar{\lambda} \leqq 1$ is chosen so that

$$\theta(\hat{x},\hat{\eta}) = \min_{0\leqq\lambda\leqq 1} \theta(\bar{x} + \lambda\bar{d}, \bar{\eta} + \lambda\bar{\delta})\ .$$

Because Theorem 3.2 and $\bar{d} \neq 0$, the number $\gamma := \theta(\bar{x},\bar{\eta}) - \theta(\hat{x},\hat{\eta})$ is positive. By Lemma 4.1 we also have $(d_k,\delta_k) \to (\bar{d},\bar{\delta})$, here again $\{(d_k,\delta_k)\}$ may be only a subsequence. Therefore, there exists an arbitrarily large k such that

$$\begin{aligned} \theta(x_{k+1},\eta_{k+1}) &= \theta(x_k + \lambda_k d_k, \eta_k + \lambda_k\delta_k) \\ &\leqq \theta(x_k + \bar{\lambda} d_k, \eta_k + \bar{\lambda}\delta_k) \\ &\leqq \theta(\hat{x},\hat{\eta}) + \frac{1}{2}\gamma \\ &< \theta(\bar{x},\bar{\eta})\ . \end{aligned}$$

This contradicts the fact that the sequence $\{\theta(x_k,\eta_k)\}$ is monoton decreasing and $\theta(x_{k+1},\eta_{k+1}) > \theta(\bar{x},\bar{\eta})$. Q.E.D.

The sequence of points can be shown to converge to a solution point when we assume that the functions f_i are convex. But, from Theorem 4.2 our method should be expected to work well even when the functions are not convex.

REFERENCES

1. J. M. Danskin, "The Theory of Max-Min, with Applications," SIAM J. Appl. Math. 14 (1966), pp. 641-664.

2. V. F. Dem'yanov and V. N. Malozemov, Introduction to Minimax, John Wiley & Sons, New York, 1974.

3. S. P. Han, "A Global Convergent Method for Nonlinear Programming," J. of Optimization Theory & Appl., 22 (1977), pp. 297-309.

4. S. P. Han, "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," Mathematical Programming, 11 (1976), pp. 263-282.

5. S. P. Han, "Variable Metric Methods for Minimizing a Class of Nondifferentiable Functions," to appear in Mathematical Programming.

6. S. P. Han and O. L. Mangasarian, "Exact Penalty Functions in Nonlinear Programming," in preparation.

7. O. L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York, 1969.

8. M. J. D. Powell, "A Fast Algorithm for Nonlinear Constrained Optimization Calculations," presented in the 1977 Dundee Conference on Numerical Analysis.

9. M. J. D. Powell, "The Convergence of Variable Metric Methods for Nonlinear Constrained Optimization Calculations," in: Nonlinear Programming 3, ed. O. L. Mangasarian, R. R. Meyer & S. M. Robinson, Academic Press, New York, 1978.

10. S. M. Robinson, "Stability Theory of Systems of Inequalities, Part II: Differentiable Nonlinear Systems," SIAM J. Numer. Anal., 13 (1976), pp. 497-513.

11. S. M. Robinson, "Generalized Equations and Their Solutions, Part I: Basic Theory," MRC Technical Summary Report #1812, 1977.

SPH/ed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1891	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON THE VALIDITY OF A NONLINEAR PROGRAMMING METHOD FOR SOLVING MINIMAX PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Shih-Ping Han		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 5 - Mathematical Programming and Operations Research
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE October 1978
		13. NUMBER OF PAGES 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Minimax, Nonlinear programming, Quasi-Newton method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the minimization of a function which is the maximum of a finite number of smooth but nonlinear functions. It is well-known that the minimax problem of this type connects naturally to a nonlinear program. Through this connection the effective quasi-Newton method becomes applicable. We show that this approach is valid and the resulting method has global convergence properties.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)